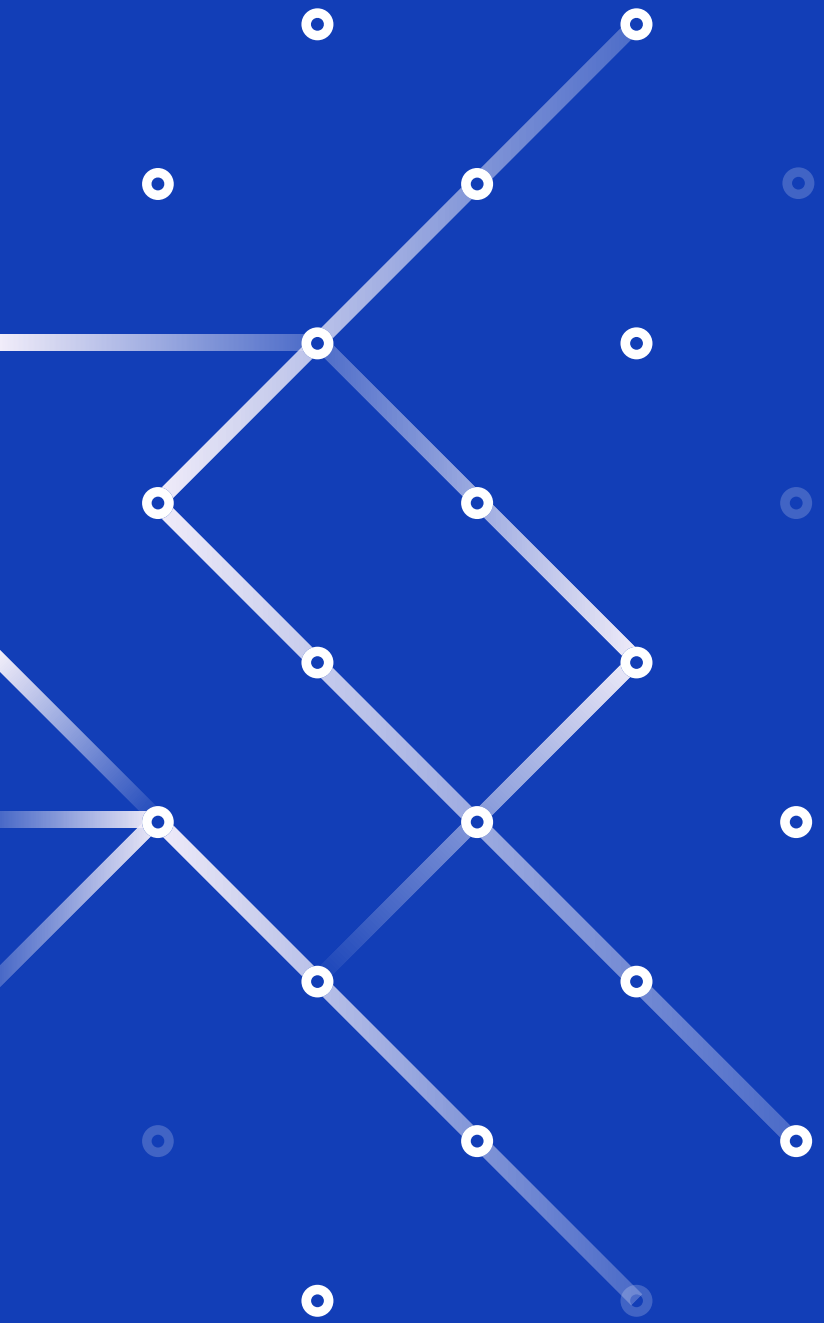


Responsible AI

The Rule of Law

Authors

Liisa Janssens LLM MA, Okke Lucassen MA, Laura Middeldorp MSc, Larissa Lobbezoo MSc, Olivier Schoenmakers MSc

July 2024

Table of Contents

Responsible AI
The Rule of Law

Authors
Liisa Janssens LLM MA, Okke Lucassen MA, Laura Middeldorp MSc, Larissa Lobbezoo MSc, Olivier Schoenmakers MSc

July 2024

Introduction

Characteristics of a constitutional society are separation of powers and checks and balances. The Rule of Law shapes the roles and mandates in order to protect society against inside and outside threats. The legislative, executive and judicial power need to function within the Rule of Law. Protection of society is key to the role of the executive power of which law enforcement and the military are both players. Means and methods to do this require adequacy, efficiency and trustworthiness.

Emerging disruptive technologies (EDTs) can both support and disrupt activities deployed by the executive power. In this paper the focus is set on law enforcement and the military, or in short: the defense, safety and security domain.
Examples of support can be found in the opportunities of –for example- biometric analysis (e.g. facial recognition), digital forensics (e.g. detection of social engineering), and online threat analysis (e.g. recognition of hate speech) (Daniel Lückerath, 2023). Despite the opportunities of EDTs, these examples can also lead to disruptive situations in civil and military operations.

Especially, it can be difficult to determine when and to what extent it is responsible in operations to use these technologies in the context of the defense, safety and security domain. Specifically, the impact of AI systems (OECD, 2023) on existing (and future) laws, rules and regulations should be assessed to establish responsible AI.

An example of (future) legislation is the AI Act (AIA). This legislation aims to regulate the internal market of the European Union. European law enforcement is subject to this regulation. Furthermore, the values of the European Union, such as the Rule of Law, are underpinned in the AIA.

Although the AIA is not yet enforced, two things are certain: the execution of the AIA is an interdisciplinary effort and this effort needs to adhere to the Rule of Law. Although defense, in military operations, is excluded from the AIA, these aspects are key to establish responsible AI in the defense, safety and security domain.

This paper presents key elements of a new method that is based on an interdisciplinary approach in which three disciplines are presented: law, philosophy of law and technology. All three disciplines, together with scenario planning, contribute to the establishment of responsible AI systems.

Responsible AI systems namely require an integrative level of understanding of what is at stake. Scenarios illustrate how the use of AI systems can (unintentionally) undermine the Rule of Law and, on the other side, how the use of AI systems can be an absolute necessity to protect society.

To these ends, TNO has developed an interdisciplinary method which is built on scientific knowledge and practical experiences retrieved from applied projects conducted in the civil and military domain (Dutch National Police and NATO).

The method leads to the formulation of functional, technical and operational requirements for specific AI systems that support responsible use in the context of defense, safety and security.

Goal

This paper describes the potential of scenarios in a new (scenario-based) method. Scenarios can function as a platform for interdisciplinary research to specify what responsible AI systems entail and at the same time inform stakeholders in the defense, safety and security domain -who have legitimized decision power- how to navigate.

The goal of our (scenario-based) method is to formulate requirements which are tangible and informed by (various) responsible AI frameworks, law and the Rule of Law; with the goal of practical implementation.

This method is formed through research that is conducted in two finished (multi year) applied projects : 'AI in Counter Unmanned Aircraft Systems and the Rule of Law' for the Dutch National Police and 'The Design of AI in C-UAS and the Rule of Law' for NATO.

These projects in the defense, safety and security domain can be seen as a starting point of applying and further developing this method. Just like the dynamic nature of AI, the method is constantly under construction and will flourish even more when new applied projects are researched via this method.

The key aspects of the method are:

1. **Capability gap analysis**
2. **Scenario design**
3. **Requirement identification**
4. **Validating, verifying and implementing requirements**

The key aspects will be linked to the Rule of Law and (governmental) institutions in society who have the obligation to adhere to the Rule of Law.

The goal of this paper is to get the reader acquainted with this method and to inform how this method can be applicable to other use cases. The more use cases are researched, the more the outcome will become valuable for society.

First, we will elaborate on Artificial Intelligence as an emerging disruptive technology (**Chapter 1**).

Second, we describe which tenets can be retrieved from the Rule of Law and explain how AI is challenging the Rule of Law (**Chapter 2**).

Third, we explain the key aspects of the scenario-based method and showcase the requirements analysis (validating, verifying and implementing requirements) (**Chapter 3**).

Lastly we show how this method can be applied to a use case, e.g. what goes in and comes out when applying the method -and discuss the way ahead.

1. Emerging disruptive technologies and Artificial Intelligence

Technological changes in our lives are predicted to be brought about by emerging disruptive technologies of which AI is seen as the greatest changemaker.

Together with blockchain, quantum and several other technology areas, AI is considered to be the most disruptive technology and has impacted numerous sectors and domains. (Păvăloaia, 2023)

EDTs have the potential to revolutionise the defense, safety and security domain. The EU and its Member States have recognised the importance of EDTs and have launched several initiatives and dedicated substantial funds to EDT Research and Development (R&D). (Clapp, 2023)

However, keeping up with adversaries in this area is not a given and attention is needed from both national and European perspective.

Beyond European context, EDTs also change the way NATO operates. In international military context EDTs present, according to NATO, both risks and opportunities:

"That is why the Alliance is working with public and private sector partners, academia and civil society to develop and adopt new technologies, establish international principles of responsible use and maintain NATO's technological edge through innovation." (NATO, 2023)

In international military context, EDTs can strengthen the Alliance's edge. NATO has identified the following emerging (disruptive) technology areas:

- Artificial intelligence (AI)
- Autonomy
- Quantum
- Biotechnologies and human enhancement
- Hypersonic systems Space
- Novel materials and manufacturing
- Energy and propulsion
- Next-generation communications networks

(NATO, 2023)

Via our scenario-based method we illustrate distinctions between the different EDTs, for example between Artificial Intelligence and Autonomy.

Furthermore the scenario-based method shows, in different applied use cases and (civil-military) operations, the interwovenness with AI and EDTs.

1.1. Definitions of AI

The security landscape is changed by EDTs. To make it even more complex: all the EDT technology areas have **interwovenness with AI.** This makes the changing security landscape challenging and unpredictable to navigate.

The scientific field of Artificial Intelligence (AI) is developing rapidly: the number of AI-related publications has grown by more than six-fold compared to 10 years ago (Nestor Maslej, 2023).

A wide variety of definitions on AI exist formulated by (inter)national organisations and governmental bodies such as NATO, OECD, EU, UNESCO, the UK MoD and the US DoD.

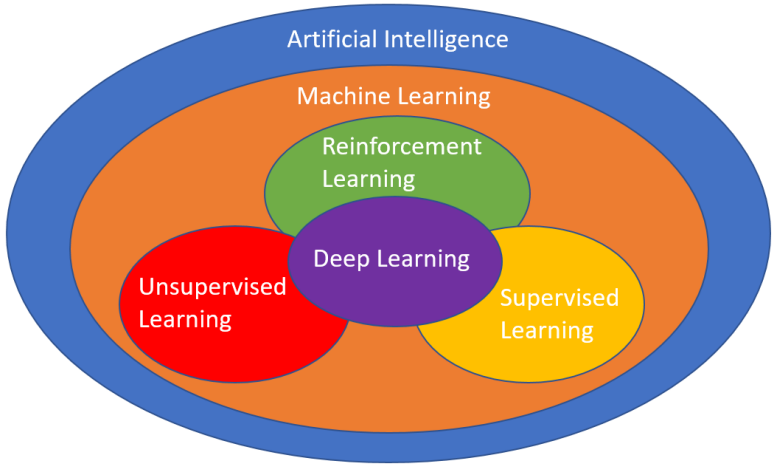
OECD defines AI as:

"An AI system is a machine-based system that is **capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives**.
It uses machine and/or human-based data and inputs to
(i) perceive real and/or virtual environments;
(ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and
(iii) use model inference to formulate options for outcomes.
AI systems are designed to operate with varying levels of autonomy." (OECD, 2023)

In our scenario-based method, finding a definition of AI is not the objective. Therefore a broad definition as stated by the OECD is sufficient. The main goal of our method is to demystify the potential of AI in existing systems or stand-alone applications by applying these in civil-military operational scenarios.

1.2. The landscape of AI

The landscape of AI can be classified into three domains: AI, Machine Learning (ML) and Deep Learning (DL). (Karthikeyan, 2022)



One of the most simple forms of AI are rule-based systems: given a set of rules, the AI attempts to generates output based on a deterministic system. Machine Learning algorithms can, given a set of features and human intervention, determine the best solution given a set of data.

Within Machine Learning a distinction can be made between supervised learning, unsupervised learning and reinforcement learning. Supervised learning entails that the dataset on which the algorithm is trained is labeled while the dataset is unlabeled in the case of unsupervised learning. (IBM)

In both cases the datasets are static, unlike reinforcement learning algorithms where the algorithm does not learn from sample data but attempts to learn from trial and error.

Deep Learning is a subset of Machine Learning where the backbone forms so-called neural networks that are capable to extract the relevant features from the dataset. (IBM)

AI exists for a wide variety of applications and can be divided into different branches such as, but not limited to:

- **Computer vision:** algorithms capable to derive information/conclusions from images or videos.
- **Natural Language Processing:** algorithms capable to understand and derive information from written or speech data.
- **Generative AI:** algorithms that are capable to generate text, images, audio or other content based on an assignment specified by the user.

AI algorithms can thus range from very transparent rule-based algorithms to complicated, intractable deep learning algorithms which have a wide variety of applications. Within our research conducted for NATO *(NATO, 2024),* various AI applications were identified in the context of Counter Unmanned Aerial Systems (C-UAS). For example, a visual UAS detection system may be enriched with a Computer Vision application to increase the effectiveness of the system.

The method presented in this paper is applicable for all domains of AI.

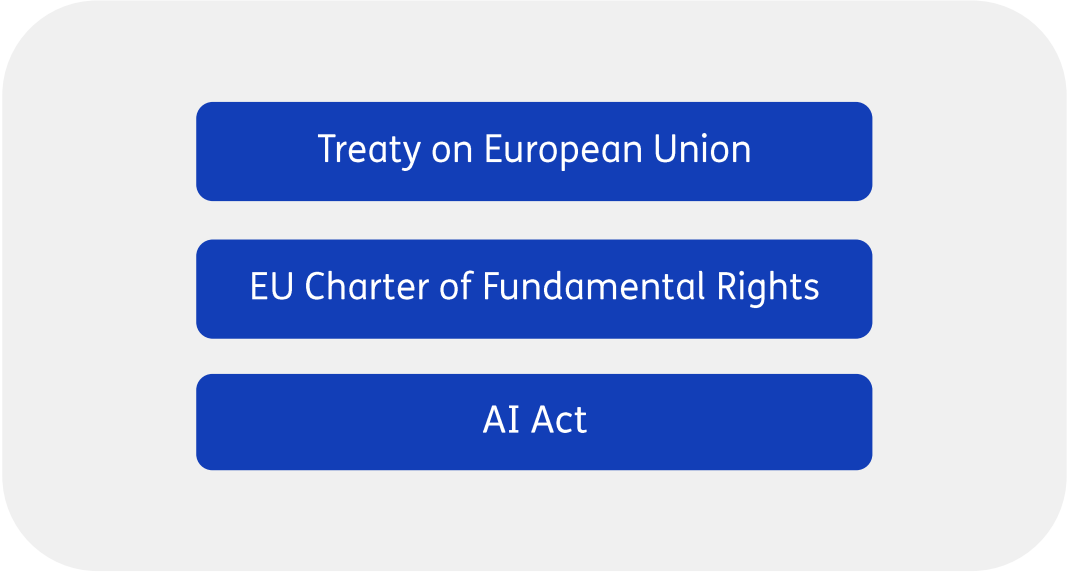
2. The Rule of Law: a European Union and international value

The Rule of Law is a value on which the European Union is founded. This makes it part of the very core of our identity. The Rule of Law needs to be guaranteed by governmental institutions, it is one of the values of article 2 of the Treaty on European Union. Values on EU need to be respected, guaranteed and promoted. The AI Act is an example of such a promotion.

The European Commission for Democracy through Law published a report on the Rule of Law, which is adopted by the Venice Commission. Herein is reaffirmed that **the Rule of Law** is one of the foundations of democratic societies as it relates to the exercise of power and the relationship between individuals and the state. It refers to the idea that the same rules, standards and principles are applied to all individuals and organisations, including government itself. **The rule of law** requires everyone to be treated in accordance with the law, with dignity, equality and rationality, and to have the opportunity of fair procedures before independent and impartial courts (Venice Commission, 2011).

Article 2 of the Treaty on European Union states that "the Union is founded on the values of respect for [...] **the rule of law** and respect for human rights". This is also described in the EU Charter of Fundamental Rights "the Union is based on the principles of democracy and **the rule of law** (...)". In the EU Accession Criteria it is stated that: "Institutions guaranteeing democracy, **the rule of law**, human rights and respect for and protection of minorities."

The AIA is a forthcoming regulation which aims to ensure that fundamental rights, democracy, **the rule of law** and environmental sustainability are protected from forbidden, high and low risk AI whilst boosting innovation. The rules in the AIA establish obligations for AI systems on its potential risks and level of impact. (European Parliament, 2023).



Sources of the European Union which underpin that the Rule of Law is part of the core identity of the European Union.

2. The Rule of Law: a European Union and international value

The Rule of Law is not only a value to foster good governance in the European Union, the Rule of Law is also underpinned in (inter)governmental institutions, entities and bodies such as OECD, NATO and the United Nations.

The OECD confirms that a multitude of statutes, laws, codes and procedures ensure that the requirements that everyone is to be treated in accordance with the law, with dignity, equality and rationality, and to have the opportunity of fair procedures before independent and impartial courts are implemented. And that "**strengthening the rule of law is considered a priority of any governance reform, as well as a key indicator of good public governance.** It is an essential prerequisite for ensuring the provision of public goods and services, economic development, maintaining peace and order, and the effective control of corruption." (Peña-López, 2019)

The importance of the Rule of Law is described in the treaty that constitutes NATO:

"The Parties to this Treaty reaffirm their faith in the purposes and principles of the Charter of the United Nations and their desire to live in peace with all peoples and all governments.

They are determined to safeguard the freedom, common heritage and civilisation of their peoples, founded on the principles of democracy, individual liberty and **the rule of law**. They seek to promote stability and well-being in the North Atlantic area.

They are resolved to **unite** their **efforts for collective defence and for the preservation of peace and security**. They therefore agree to this North Atlantic Treaty."
(NATO Treaty, 1949)

Inter-governmental organisations which underpin the importance of the Rule of Law.

2. The Rule of Law: a European Union and international value

The Rule of Law is not only a value in the European Union, the Rule of Law is also underpinned and recognized in other (inter)governmental institutions, entities and bodies such as OECD, NATO and the United Nations.

The system of the United Nations (UN) connects to the Rule of Law. The UN describes **the rule of law** as "a principle of governance in which all persons, institutions and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with international human rights norms and standards. **It requires measures to ensure adherence to the principles of supremacy of the law, equality before the law, accountability to the law, fairness in the application of the law, separation of powers, participation in decision-making, legal certainty, avoidance of arbitrariness, and procedural and legal transparency.**". (United Nations, 2024)

Inter-governmental organisations which underpin the importance of the Rule of Law.

2.1. Tenets of the Rule of Law

The Rule of Law tenets are not categorised as 'positive law'. Positive law contains rules and regulations that concerns compliance. However, the Rule of Law -and connected mechanisms- differ from principles of law that can be applied *directly* to real life use cases and thus scenarios. Existing Rule of Law mechanisms can be found in, for example, processes of new legislation, or redefining the policies and processes within the boundaries set by existing laws. (Janssens, 2023)

The Rule of Law states that all individuals and organisations, including law-makers, are accountable to the same set of laws. Furthermore, the Rule of Law is the core value of the European Union and fosters checks and balances.

Checks and balances in the context of AI systems imply that the trustworthiness of AI should be controllable and that violations can be addressed.

Our method seeks for requirements that strengthen the Rule of Law. This method fosters guaranteeing the adherence to the Rule of Law by (governmental) institutions, entities and bodies.

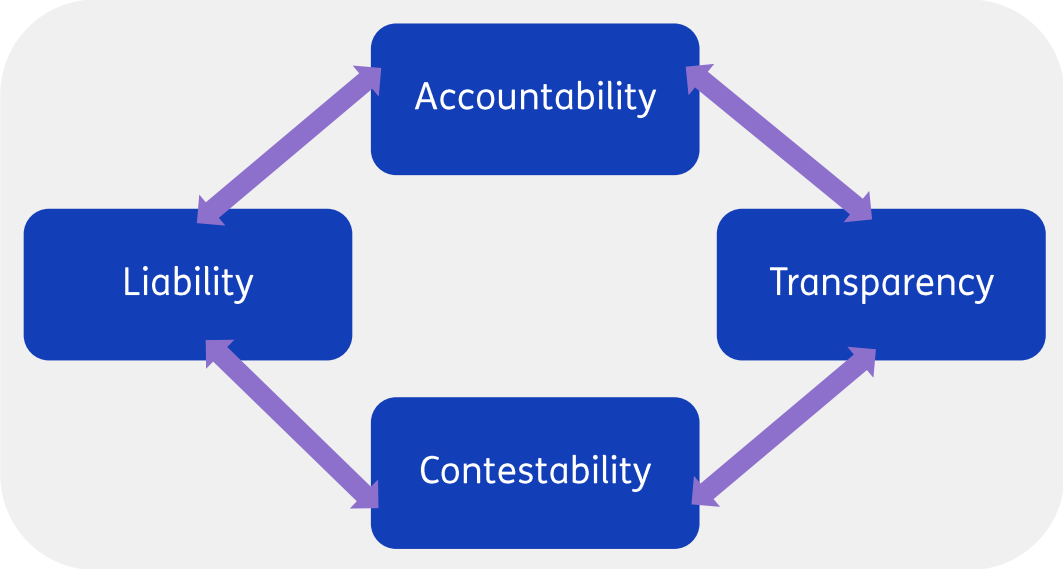
The Rule of Law has **tenets and mechanisms** to make checks and balances possible.

Examples of tenets of the Rule of Law are:

- **Accountability**,
- **Transparency**,
- **Contestability** and
- **Liability**.

The Rule of Law is shaped by many sources, such as: case law; legal doctrine; legal interpretation methods; positive law; rules and regulations; draft rules and regulations and legal theory. (Janssens, 2023)

It is important to take into account that, next to these tenets, aspects of the Rule of Law such as the **separation of powers** and **checks and balances** are taken into account in our scenario-based method.



Examples of tenets of the Rule of Law; these tenets are interrelated. Requirements for AI systems should facilitate these tenets.

2.1. Tenets of the Rule of Law

The operational use of AI systems can be challenging to the Rule of Law, since the consequences in real life setting are often unclear. Scenarios can clarify this challenge and requirements can be retrieved via the scenario-based method. Requirements are needed to benefit from opportunities while mitigating risks of harming the Rule of Law. To mitigate risks in the use of AI, tensions between AI and the Rule of Law should be identified and translated into requirements which have institutional meaning.

New requirements can inform existing **mechanisms** via (re)shaping legislation, revisiting old legal concepts anew and formulating (new) policies to ensure that the Rule of Law is maintained.

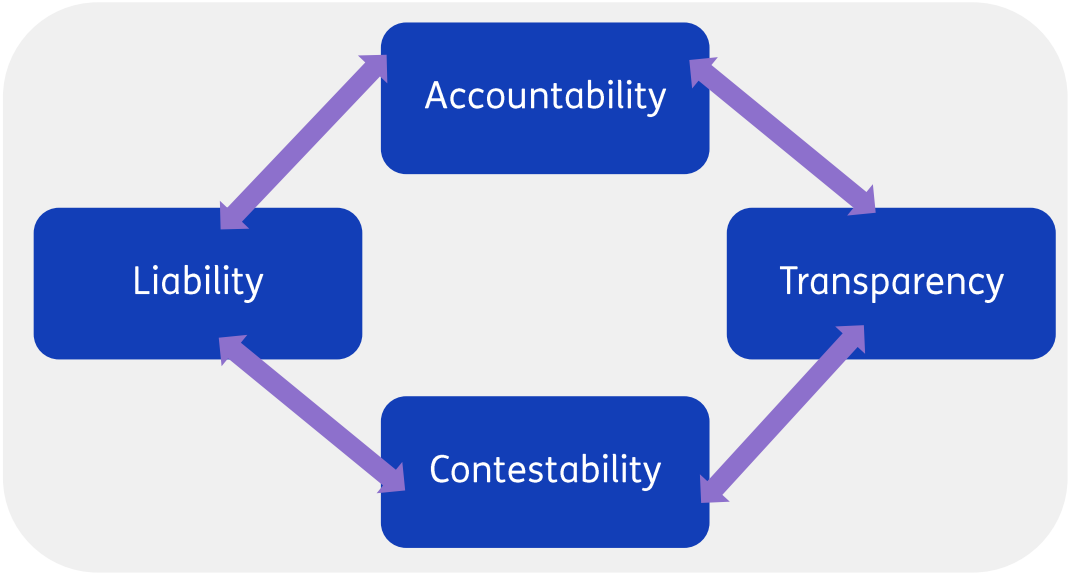
Requirements are needed to specify what responsible AI entails in the defense, safety and security domain. These requirements can have institutional meaning when these are (re)shaping laws, policies and processes.

The formulation of requirements for responsible use of AI is in this scenario-based approach an effort that comes from perspectives of law, philosophy of law and technology. Via these disciplines the impact of AI systems on fundamental values, rules and regulations is assessed.

The provisional agreement on the European AI Act (AIA) is a major accomplishment in Europe's efforts to ensure trustworthy AI (European Parliament, 2023).

Before the AIA and other (new) regulations, policies and processes are effective, the impact of AI on fundamental values, such as the Rule of Law, can be assessed via the method in order helping shape tangible requirements.

Furthermore, the assessment via the method gives illustrations on the how, when and where AI systems can protect society against internal and external threats.



Examples of tenets of the Rule of Law; these tenets are interrelated. Requirements for AI systems should facilitate these tenets.

2.2. AI is challenging the Rule of Law

So, how does AI challenge these tenets of the Rule of Law? And why are these tenets so important? A view examples in a nutshell…

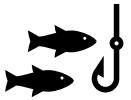
AI is an EDT that changes the concept of how to perform certain tasks compared to current ways (Clapp, 2022). Therefore, consequences of using AI are uncertain as well as the adherence to the Rule of Law. By identifying tensions with tenets of the Rule of Law, risks of using AI systems can be identified.

The tenet of **accountability** is challenged by the level of autonomy of the AI system. Suppose an AI system is fully autonomous and does not allow any human intervention, and the system makes a wrong prediction with a harmful outcome. Who or what can be held accountable? How and under which requirements can an AI system be trusted?

The tenet of **transparency** is for example challenged by AI systems that are considered a *black-box*. Most end-users do not (or cannot) understand what an answer or recommendation is based on as in-depth technical knowledge can be required. Is it safe to use these systems when there is no transparency in a R&D or procurement-process on how the AI came to its output?

The tenet of **contestability** is challenged by (unintended) misuse of data or statistics in AI systems, such as data fishing (probability-hacking) during R&D and procurement-processes. Probability hacking entails activities that modify the dataset on which the AI system is trained in such a way that a sufficient performance metric is obtained. However, the AI system performs significantly worse on a new dataset.

This can result in coincidental rather than real correlations



The tenet of **liability** is challenged by AI systems when its outcomes are poor and potentially harmful. For example, an AI system is highly dependent on its data. If the data has poor quality, the performance of the system will decrease: *rubbish in, rubbish out*. How can requirements address liability when harmful outcomes occur?

The perspective of the Rule of Law reveals potential tensions when AI is deployed by the executive power in the defense, safety and security domain.

If an AI system violates the Rule of Law, the Rule of Law is weakened. For the executive power this can mean that tasks executed with such an AI system will be in tension with norms and values that are obliged to be aligned with.

3. Scenario-based method to identify requirements

Solutions are needed to identify the tension between the Rule of Law and AI systems. Tensions have to be translated to requirements which can be institutionalized in the value chain of partners.

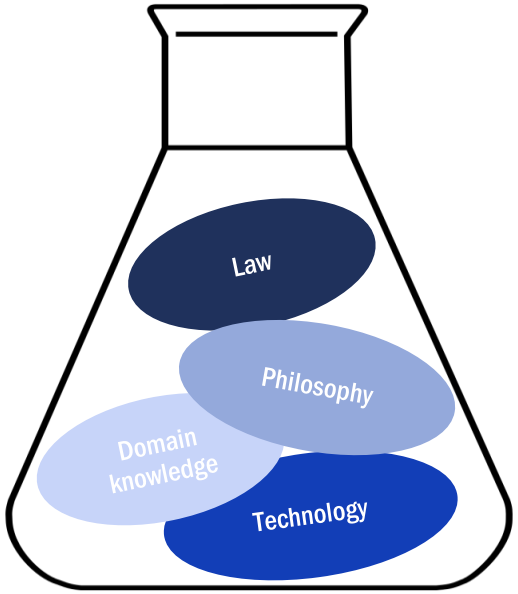
How can society be strengthened by the opportunities of AI while protecting the Rule of Law against the drawbacks of AI? The European Commission recently published the AI Act: a legal framework guiding the development and application of AI systems. This framework also applies (with some exemptions) to law enforcement operations. Aim of the AIA is upholding the values of the European Union and, at the same time, leveraging the capabilities of AI.

AI systems for the defense, safety and security domain can be beneficial as well as severe, a method is needed to identify the (unintended) consequences as specific as possible and has an effect on the whole value chain of partners who bear responsibility for the AI application from design to deployment. In order to do so our interdisciplinary team combines backgrounds in law, philosophy of law and technology with specific (technology and AI-systems driven) use-cases in scenarios.

Scenarios are a useful instrument to simulate a specific threat environment through which various AI systems can be deployed. Scenarios allow to assess the result of the use of a specific technology in a specific context over time. By mapping this to (tenets of) the Rule of Law requirements are deducted.

These requirements can be used to enhance the mechanisms of the Rule of Law via (re)shaping legislation, revisiting old legal concepts anew and formulating (new) policies to ensure that the Rule of Law is maintained.

A practical example where requirements can be implemented is introducing them into R&D and procurement-processes. Furthermore, these requirements can be linked to the activities of the European Artificial Intelligence Office which will be established (see the decision of the European Commission on the 24th of January 2024 on establishing the European Artificial Intelligence Office).



Scenarios, which are fueled by domain knowledge, are a useful instrument to bridge different disciplines and become specific on the problems and requirements to mitigate risks of harming the Rule of Law.

3.1. Trust and the system of law

Why does the method identify requirements?

The reason for the separation of powers into a legislative, executive and judicial power is enabling a system of checks and balances, where each power checks and balances the other power. Requirements can be seen as checks that AI systems should satisfy to balance power in this system of checks and balances. By adhering to the Rule of Law, the AI system promotes the balancing of power.

At the same time these requirements can lead to better functioning AI systems because the requirements enhance the **quality** and herewith the **trustworthiness** of these systems.

The method has been applied in research projects for NATO and the Dutch National Police, with a specific focus on **enhancing** Counter Unmanned Aircraft Systems (C-UAS) with AI.

Besides enhancing capabilities, requirements also play an important role in the given that trust and the system of law are complementary to each other, but cannot be replaced on for the other.

In the defense, safety and security domain trust is essential. For example in military operations commanders and mission intent are based on trust. The opportunities of speeding up decision-making in the military domain can contribute to a military advantage, but danger resides in the processes before deployment. For example, via errors in the research design of AI applications which can lead to untrustworthy accuracy rates. These untrustworthy rates can lead to collateral damage, or mistakes which in turn can lead to military disadvantages. (Janssens, 2023)

These disadvantages will not be different for operations with AI systems in law enforcement.

From a legal perspective the Rule of Law tenet **contestability** is at stake when there are no requirements that check on trustworthiness. A commander needs to be able to **contest** if a mission goes terribly wrong and he or she is accused not following the rules of engagement. The same is to say for a law enforcement officer who needs to be able to contest an accusation of ‘not doing the right thing’ when he or she suspects malfunctioning of an AI system.

The possibility to contest such an accusation is challenged when there are no additional requirements implemented in existing Rule of Law mechanisms such as R&D and procurement to prevent (unintended) misuse of data or statistics in AI systems, such as data fishing (p-hacking). This can result in coincidental rather than real correlations. It needs no explanation that this will have major negative consequences for deploying tasks in the defense, safety and security domain within the boundaries set by the Rule of Law.

3.2. Key aspects of the method

Requirements are identified for a specific AI system in a specific scenario. Requirements are derived using the tenets of the Rule of Law, responsible AI frameworks and technical paradigms.

1. Capability gap analysis

The first step is to perform a capability gap analysis: which existing problems can be solved or mitigated with AI? The problem may be linked to a system, e.g., a radar system, but can also be a stand-alone problem.

In the first case the system is enriched with AI, while in the second case the AI itself constitutes the solution in the form as a stand alone application. Given the problem, literature needs to be reviewed to identify the potential of various types of AI corresponding to a system or problem. In our research on Counter Unmanned Aircraft Systems (C-UAS), a database has been developed. The capability gap analysis can also be applied in other use cases.

The database distinguishes three types of C-UAS with regard to AI:

1. C-UAS for which no AI applications have been identified yet;
2. C-UAS for which AI applications are present in literature;
3. C-UAS with potential for adding AI applications that are not yet present in literature.

An example of an available C-UAS with AI software are the IRIS® or ELVIRA® drone detection radars which can automatically classify drones using ML-based software to support early-warning (Robin Radar, 2023).

Because the amount of scientific research on AI is rapidly growing, it is important to keep track of all publications related to AI and specific systems. Furthermore, because of this constantly extending research, the AI for a specific system database requires updates when new AI applications arise.

2. Scenario design

The second step is to formulate a scenario in which AI plays a role. The scenario describes the context in which AI is operationalised and consists of several factors that fuel into a scenario, such as the legal basis and legal regime, threat level, weather, landscape and actors (e.g. end-users, civilians, malicious parties).

A key element of the use of scenarios is that these can be constructed on several levels: strategic, operational, tactical and technical. The level of the scenario is related to which stakeholders you need to involve and inform how, and what type of, requirements can establish responsible AI.

Building and shaping scenarios is an ongoing iterative process.

3. Requirement identification (via A, B and C)

The third step is to analyse the scenario with the specific problem, linked to operational practice, via a list of requirements. Since there are many (policy) frameworks from (governmental, non-governmental and standardisation) institutions, entities and bodies on responsible AI, the deduction of a requirements-list depends on the desired framework for responsible AI that needs to be analysed and contextualised via the Rule of Law.

3A Frameworks for Responsible AI

Several (inter)governmental, defence and standardisation organisations, such as NATO (Zoe Stanley-Lockman, 2021), OECD, ISO, CEN and IEEE have developed their own frameworks for AI. The ambition of these frameworks is to provide guidelines for responsible AI. These guidelines are given by principles which describe aspects that the AI system should adhere to. It is important to know how to select applicable frameworks for specific operations.

For example, a framework can give guidelines about how the dataset used to train the AI system should be free of bias. However, the translation from principle to responsible practice is often missing. Furthermore the institutional meaning if often unclear. What does responsible AI entail in operations and how can these frameworks have institutional meaning? First: whilst not all the operational consequences can be overseen prior to deployment a contextualisation of these principles can illustrate what is at stake. Second: a list of requirements need to be informed by this contextualisation and stakeholders with decision power need to be informed. The list of requirements is linked how principles from these guidelines can have institutional meaning via informing specific existing laws, policies and processes.

Next to examples of R&D or procurement-processes, on EU level the European AI Office will be an important vehicle. In our method, the most applicable frameworks are chosen based on the stakeholder. For instance, if the stakeholder is from the Ministry of Defence, the NATO's AI Strategy and framework of Principles of Responsible Use is applicable, while a framework such as the OECD AI Principles overview is more suitable for a national safety and security stakeholder, such as the Dutch National Police or Custodial Institutions Agency (DJI).

3B Technical paradigms

The requirements are formulated by mapping the principles of the chosen framework to the principles retrieved from a chosen technical paradigm, for example that of Machine Learning Operations (MLOps) (Larysa Visengeriyeva, 2023).

MLOps, a compound of Machine Learning (ML) and IT Operations (Ops), is a set of practices with the aim to deploy and maintain Machine Learning (ML) systems in production reliably and efficiently.

The principles of MLOps are informed by technology and constitute versioning, testing, automation, reproducibility, deployment and monitoring.

3C Viewed through the lens of the Rule of Law

The requirements are a reflection of norms and values in our society framed in checks on a technical AI system. Ideally, the requirements list will reflect the proper functioning of the tenets of the Rule of Law and benefits from technical principles.

While a technical paradigm (such as MLOps) focuses on the technical aspects of requirements for AI, the principles of the chosen framework link to how this can be done in a responsible way throughout the value chain of partners. When these together are viewed through the lens of the Rule of Law you have 'the best of both worlds' in formulating requirements.

4. Validating, verifying and implementing requirements

One of the objectives of the method is to take the first step in **operationalising** principles of frameworks via the means of a scenario.

First, preselected **requirements** can be validated and verified for AI in the specific scenario by asking the following question: how does AI in the scenario satisfy a requirement and if not, how can it be ensured that it does?

Second, **new requirements** can be identified based on the scenario which are not yet incorporated in the requirements list. Note that these requirements are scenario-specific and are not necessarily valid when the AI is placed in a different scenario.

A **feasibility test** of (*newly found*) requirements is still subject of further research, and we are exploring how this can be done via table-top workshops, simulation or operational tests.

Overall, the requirements-list can be seen as a checklist that must be taken into account from development to deployment throughout the value chain of partners. Therefore, the requirements-list acts as a protective agent for respecting, guaranteeing and promoting the Rule of Law.

This operationalisation can inform, in order to deploy AI systems in a responsible way, stakeholders (in the value chain of partners) with legitimized decision power about the necessity of implementing new requirements in existing Rule of Law mechanisms.

3.3. Input and output

So, what goes in and what comes out of the scenario-based method?

The inputs for the method are:

- **a specific problem**: also called a **capability gap**, for which an AI system is identified;
- **a specific scenario**: the context of the AI system is given by a scenario that describes which factors and actors influence or intervene with the AI system;
- **a specific framework for responsible AI:** to formulate requirements corresponding to the principles;
- **principles** from a technical paradigm (for example **MLOps**): to formulate functional, operational and technical requirements;
- the **Rule of Law**: and in its interrelation with case law; legal doctrine; legal interpretation methods; positive law; rules and regulations; draft rules and regulations and legal theory.

The outputs of the method are:

- a **validation and verification** of existing institutionalised requirements;
- **identification** of **new requirements** which can be institutionalised via Rule of Law mechanisms.

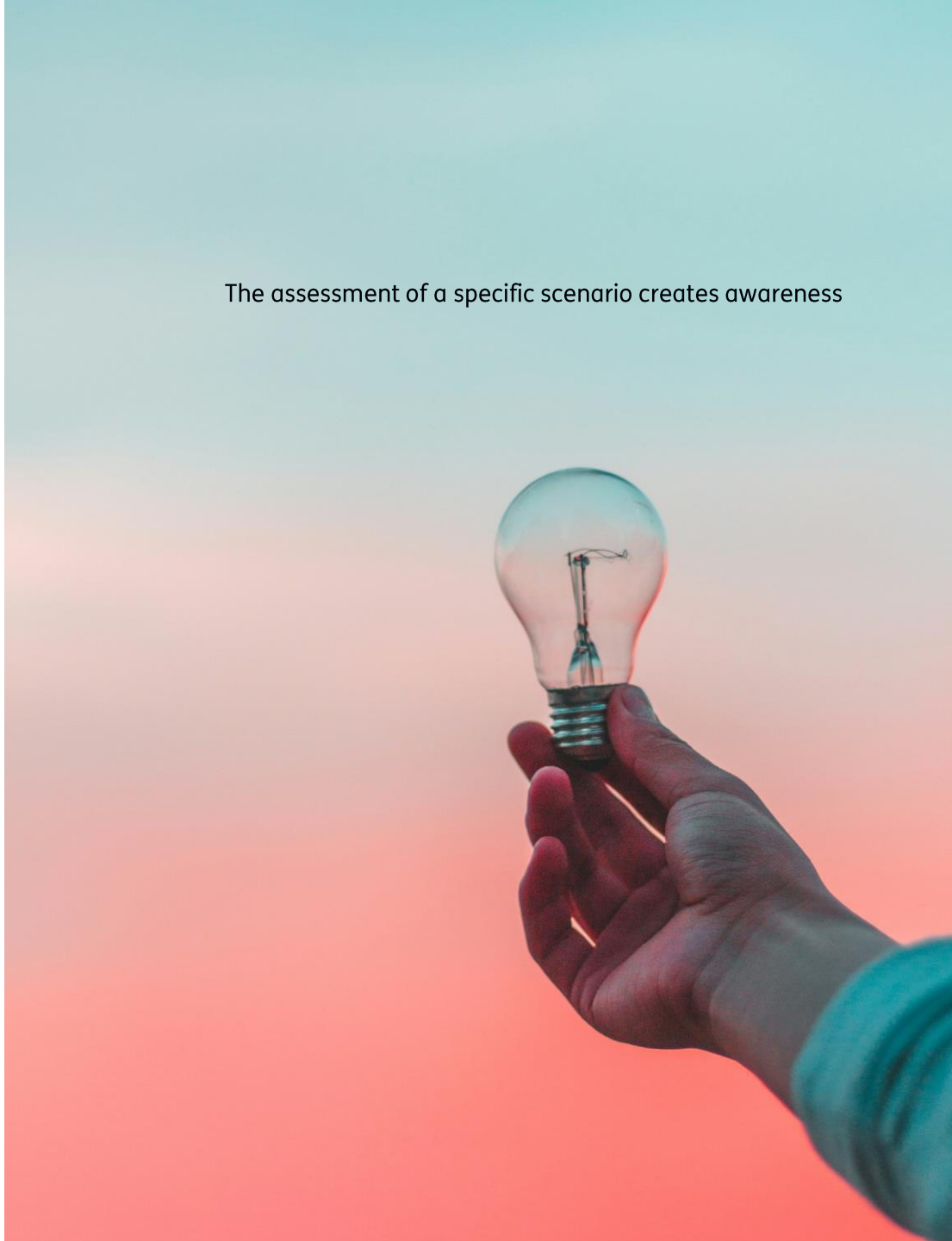
What can be achieved with the output?

The formulated requirements can have institutional meaning via (re)shaping legislation, revisiting old legal concepts anew and formulating (new) policies to ensure that the Rule of Law is maintained, and at the same time enhancing systems with AI towards establishing problem solutions which are adequate, efficient and trustworthy.

Implementation of these requirements can be done (for example) via:

- **Research & Development** processes;
- **Procurement** processes;
- European AI Office activities.

The assessment of a specific scenario creates awareness



Way ahead

Summarizing, this paper described a scenario-based method for responsible AI. Because of the constant development of AI this method will constantly be under construction, wherein scenarios are used as an interdisciplinary platform. Via this platform we are able to seek for an integrative level of understanding of what it entails to establish responsible AI.

This method can not only be helpful to become more specific on the meaning of responsible AI it also provides applied knowledge for legitimized decision makers to enable them to protect what is at stake. On the one hand this scenario-based method illustrates what is at stake when applying AI systems in law enforcement and military operations, and on the other hand the role of AI systems in protecting what is at stake becomes clear.

Applying this method helps setting up requirements which can (re)shape existing laws, policies and processes. Via these Rule of Law mechanisms, stakeholders (which have legitimized decision power) can implement the new identified requirements. After implementation in Rule of Law mechanisms the requirements will have institutional meaning.

Although the scenario-based method will be constantly under development, the following can already be taken from this:

- **Awareness** of possible violations of the Rule of Law, when the AI system is used in an operational setting, is necessary to protect the Rule of Law.
- **Scenarios can be a tool** to formulate interdisciplinary informed requirements. In the current state, as formulated in this paper, it can be seen as a thought experiment to translate norms, standards and values into requirements that can have institutional meaning through the lens of the Rule of Law.
- **Feasibility tests** of the *newly found* requirements is still subject of further research, since we are exploring how this can be done fruitfully via table-top workshops, simulation or operational tests.
- **Executing responsible AI** within the defense, safety and security domain will become an interdisciplinary effort, and this effort needs to adhere to the Rule of Law, this scenario-based method provides an answer to that need.

Dutch National Police

The research conducted for the Dutch National Police has focused on how AI applications for C-UAS can be responsibly procured within the safety and security context. For the Dutch National Police it is important to innovate. Since innovation can contribute to efficient use of (public) resources and enhancing executing tasks as set in the Politiewet 2012.

Innovations to counter drones more adequately can be done via AI applications, especially in the field of C-UAS since UAS are becoming more prevalent and hence form a higher risk within the safety and security domain.

However, at the time of the research, from January 2021 until January 2023, the AI Act of the European Commission had not been published yet. Therefore, no concrete policy or legislation was in place that specifically monitored if the development, procurement and use of AI applications were performed in a responsible manner. The Dutch National Police had no available tools to determine if an AI application is used responsibly and aligned with the Rule of Law. Our research has provided the Dutch National Police with a methodology that can be used to answer these questions.

A database has been constructed that lists possible AI applications for C-UAS in the safety and security setting. This database has been used to formulate a scenario in which the following three C-UAS, enriched with AI applications, have been examined:

1. AI application used to strengthen the information position of the Dutch National Police on possible threats.

   Natural Language Processing (NLP) is applied on Open Source Intelligence (OSINT) to perform social media monitoring. Computer vision is used to analyse satellite images to discover placements of possibly dangerous materials.

2. AI application used to identify suspicious individuals.

   Given camera footage, the AI application detects suspicious behavior based on the movements of individuals. If a person is deemed suspicious, he/she is being traced and all movements are recorded.

3. AI application that predicts the flight path of a UAS based on the coordinates and type of the UAS.

   The predicted flight path is given as input to a Radio Frequency (RF) jamming system.

The contents of the scenario are classified and therefore not shared in this paper.

Grosso modo: for each AI application, within the context of the scenario, has been evaluated on risks, especially on harming the Rule of Law on the following aspects:

- Legality and legitimacy;
- Accountability to the law and contestability;
- The link with the AI Act.

After scoping risks, requirements -which have the aim to mitigate these risks- are formulated.

NATO

NATO has officially adopted an Artificial Intelligence strategy in October 2021. The revised AI strategy is published on the 10th of July 2024 and comprised of six principles:

- Lawfulness;
- Responsibility and accountability;
- Explainability and traceability;
- Reliability;
- Governability;
- Bias mitigation.

By committing to the, NATO and allies ensure that the AI applications they develop and deploy are in accordance with the six principles. The main question is how these principles translate to AI applications being used in practice. In other words: how can the principles be operationalized? Within our NATO research, we have applied the scenario-based method to answer this question. Although our method is applicable to all means and methods of warfare which deal with emerging disruptive technologies, in the NATO report 'The Design of AI and the Rule of Law' we have specifically focused on AI applications for C-UAS.

First, a capability gap analysis has been performed to identify problems within C-UAS that may be solved with AI. For each capability gap, a literature study has been carried out to find potential AI applications that may fill the capability gap. The capability gaps together with the potential AI applications form the database of AI applications for C-UAS.

We have chosen the MLOps framework as technical paradigm. This has been mapped on the NATO principles to obtain a list of requirements: each NATO principle has a separate list of requirements. An AI application is deemed responsibly procured/ developed if all requirements for all principles are satisfied.

Three capability gaps have been highlighted within the NATO research. For each capability gap, a potential AI application is mentioned and the requirements are tested using a scenario.

Each capability gap and AI application gets a separate scenario. The following blocks briefly describe the examined capability gaps and AI applications.

1. Capability gap: Detection of contradictory information present in a large number of documents.

   Potential AI application: Natural Language Processing (NLP) techniques.

2. Capability gap: Detection of UAS with a small Radar Cross Section (RCS) at a low altitude.

   Potential AI application: Recurrent Neural Network (RNN) to improve detection of UAS at low altitude having a small RCS.

3. Capability gap: Obtain a reliable estimate of the payload of a UAS at a long distance.

   Potential AI application: Naïve Bayes model to recognize the size of the payload of the UAS using Micro-Doppler Signatures (MDS).

Bibliography

Picture on '3.3 Input and output' copy right free: thanks to https://unsplash.com/@jdiegoph

Clapp, S. (2022). *Emerging disruptive technologies in defence*. European Parliamentary Research Service.

Daniel Lückerath, V. W. (2023). *First Update of the Research Roadmap for AI in Support of Law Enforcement and Policing*. Sankt Augustin: ALIGNER Project.

European Parliament. (2023, 12 9). *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*. Opgehaald van European Parliament: https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy AI*. Brussels: European Commission.

IBM. *What is Artificial Intelligence (AI)?* What is Artificial Intelligence (AI)? | IBM

Janssens, L.A.W. (2023). *Clarifying Military Advantages and Risks of AI Applications via a Scenario*, in preliminary book publication: The Quest for AI Sovereignty, Transparency and Accountability Preliminary version presented at the United Nations Internet Governance Forum 2023, Kyoto, Japan.

Karthikeyan, A. &. (2022). Artificial intelligence: machine learning for chemical sciences. Journal of Chemical Sciences, 134, 1-20.

Larysa Visengeriyeva, A. K. (2023). *MLOps Principles. Machine Learning Operations* Retrieved in 2023. accessed: https://ml-ops.org/content/mlops-principles

NATO Treaty, The North Atlantic Treaty Washington D.C. - 4 April 1949

NATO (2023, June 22). *Emerging and disruptive technologies*. (NATO) Retrieved August 16, 2023, from https://www.nato.int/cps/en/natohq/topics_184303.htm

NATO (2024), Janssens L.A.W., Lucassen O., Middeldorp L., Lobbezoo, L.. Forthcoming report: The Design of AI in C-UAS and the Rule of Law.

Nestor Maslej, L. F. (2023). *The AI Index 2023 Annual Report*. Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

OECD. AI Policy Observatory. (2023). *OECD AI Principles overview*. Retrieved August 17, 2023, from https://oecd.ai/en/ai-principles

Păvăloaia, V. D., & Necula, S. C. (2023). *Artificial intelligence as a disruptive technology—a systematic literature review. Electronics*, *12*(5), 1102.

Peña-López, I. (2019). Government at a Glance 2019.

Report on the rule of law - Adopted by the Venice Commission at its 86th plenary session (Venice, 25-26 March 2011) Study No. 512 / 2009 CDL-AD(2011)003rev-e

Robin Radar. (2023). *IRIS® Counter-Drone Radar*. Opgehaald van Robin Radar: https://www.robinradar.com/iris-counter-drone-radar

United Nations and the Rule of Law (2024). What is the Rule of Law, Retrieved January 25, 2024, accessed: https://www.un.org/ruleoflaw/what-is-the-rule-of-law/

Zoe Stanley-Lockman, E. H. (2021, 10 25). *An Artificial Intelligence Strategy for NATO*. Opgehaald van NATO: https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html

Contact
Liisa Janssens LLM MA
liisa.janssens@tno.nl

Responsible AI
The Rule of Law

Authors
Liisa Janssens LLM MA, Okke Lucassen MA, Laura Middeldorp MSc, Larissa Lobbezoo MSc, Olivier Schoenmakers MSc

July 2024